# 録音位置の違いによる音声信号のずれ検知

1200304　尾辻 明里　【 ネットワーク信号処理研究室 】

## 1　はじめに

近年，音声認識技術は急速に発展を遂げ，会議での議事録作成や映像の字幕付与などに貢献している．実際にスマートフォンなどで動作しているものの性能は理想的な環境では極めて高い．音声認識のシステムが用いられる会議などの実環境では，音源数が複数であるため，個々の発話の認識率を高める必要がある．そこで，所望音のみを抽出する技術が必要である．本研究では，録音された音声の中から所望音以外の音を取り除くことで所望音を抽出する方法を提案する．

## 2　所望音の抽出方法

録音された音声から所望音を抽出するには，所望音以外の音を減算できればよい．図 1 のような 2 本のマイクを用いたモデルを考える．音声 1 を所望音とすると，観測信号 1 から音声 2 の信号を除くことで所望音を抽出できる．今観測信号 1 に含まれている音声 2 の信号はもとの信号とずれている．そのずれを検知する必要があるため，所望音以外の音を消すには観測信号の完全な同期が求められる．信号のずれは，振幅の減衰や時刻のずれなど様々な要因によって生じている．今回はマイクに到達する時刻のずれに着目し，ずれを一致させる．

## 3　マイクに到達する時刻のずれ

マイクに到達する時刻のずれ $s$ サンプルは

$$s = \quad fs \times \frac{d}{c} \tag{1}$$

より予測されるはずである．ここで，$fs$ はサンプリング周波数，$c$ は音速，$d$ はマイク間の距離である．この予測が実際のずれと一致しているか実験で検証する．

### 3.1　予測したずれの誤差を求める実験

予測されるずれが実際のずれと一致しているか測定し，検証する．図 2 のようにスピーカーから 7[cm] 離れたマイク 1 と，マイク 1 から $d(20, 21, \cdots, 25)$[cm] 離れたマイク 2 を一直線上に配置し，音声信号を記録した．予測されるずれは式 (1) を用いて求め，実際のずれは相互相関関数を用いて求める．

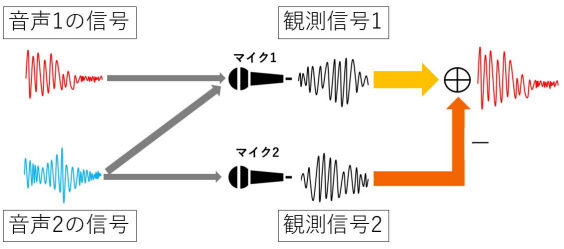


図 1　所望音の抽出のモデル

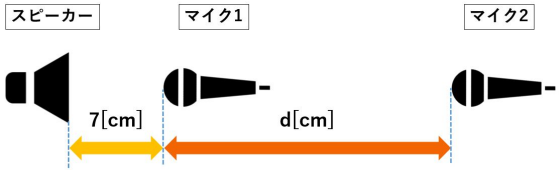


図 2　実験環境

表 1　距離毎のずれの比較

| $d$[cm] | 20 | 21 | 22 | 23 | 24 | 25 |
|---|---|---|---|---|---|---|
| $s$ サンプル | 25 | 26 | 28 | 29 | 30 | 32 |
| $n$ サンプル | 23 | 24 | 26 | 26 | 28 | 29 |
| $r_{xy}(n)$ | 0.901 | 0.897 | 0.915 | 0.924 | 0.920 | 0.917 |

### 3.2　実験結果

表 1 に実験の結果を示す．各 $d$ の距離に対して，$s$ が予測されるずれ，$n$ が実際のずれ，$r_{xy}(n)$ が相関係数を示している．表 1 より，$s$ と $n$ は一致せず，2，3 サンプルの差であった．また，$r_{xy}(n)$ は最大で 0.924 となり 1 とはならず，信号は完全に一致しなかった．これは，信号がマイクに到達するまでに減衰や残響などの影響で信号がずれたのだと考えられる．そこで，相互相関で位相のずれを補正した信号と実際に観測された信号との誤差を適応フィルタで最小化し，信号の一致を目指す [1]．

## 4　相互相関と適応フィルタを用いた所望音抽出

位相以外のずれは適応フィルタで補正し，観測信号の同期を図る．マイク 1 の観測信号を $x(n)$，マイク 2 の観測信号を $y(n)$ とする．このとき，$x(n)$ と $y(n)$ は $s$ サンプルずれているとすると，出力信号は $a(n)x(n+s)$ となる．所望信号 $y(n)$ と出力信号 $a(n)x(n+s)$ の誤差信号 $e(n)$ の 2 乗平均値 $E[e^2(n)]$ が最小になるようにフィルタ係数 $a(n)$ を更新し，誤差を最小にしていく．適応フィルタを用いた結果，マイク 1 とマイク 2 の観測信号の誤差は $1 \times 10^{-17}$ 未満であり，信号が一致することを確認した．

## 5　まとめ

本研究では，所望音以外の音を取り除くことで所望音を抽出する方法を提案した．提案方法により，録音位置の違う音声信号が一致し，所望音を抽出できることを確認した．今後は，スピーカーとマイクの配置が一直線上でない場合にも提案方法が可能であることを確認する必要がある．

## 参考文献

[1] 辻井重男，“適応信号処理，” 昭晃堂，1995.